DOCUMENT RESUME

ED 237 547 TM 830 809

AUTHOR Ferguson, William F.
TITLE Non-Traditional Answer-Sheet Format: Solution or Problem?
PUB DATE Nov 83
NOTE 13p.; Paper presented at the Annual Meeting of the Mid-South Educational Research Association (12th, Nashville, TN, November 16-18, 1983).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Answer Sheets; *Cheating; Higher Education; *Multiple Choice Tests; *Test Construction; *Test Format; Testing; Testing Problems
IDENTIFIERS Test Score Variance; Test Security

ABSTRACT

College undergraduates (n=38) were administered identical multiple choice tests with randomly presented answer-sheets numbered either vertically or horizontally. Of the originally-scheduled four tests during the semester, tests one and three were retested with entirely different test questions, also multiple choice, resulting in scores from tests, 1A, 1B, 2, 3A, 3B, and 4. Test 1B was mentioned as a possibility prior to testing of 1A. The instructor was vague and noncommittal about the possibility of a test 2A and a 2B. Prior to test 3A, subjects were told they would be retested on the same material. The students were told very definitely that test 4 would be given only once. Therefore, each of the four major content divisions of the course had tests administered under four completely different instructions. These conditions seemed to play a role in the final outcome of students' scores. When students were tested a second time without advance notice, their second mean score was lower than their first mean score, reflecting the "memory (forgetting) curve." There was a consistent pattern of a reduction of variance in the scores of the horizontally-numbered answer-sheet group. Students who performed best in the course academically did slightly better on the vertically-numbered forms. Conversely, students who did the worst academically actually did considerably better on the horizontally-numbered answer-sheet than on the vertically-numbered one. The conclusion is that the non-traditional answer-sheet format was somewhat confusing to the students. (PN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED237547

NON-TRADITIONAL ANSWER-SHEET FORMAT: SOLUTION OR PROBLEM?

MID-SOUTH EDUCATIONAL RESEARCH ASSOCIATION
Fall, 1983

by

William F. Ferguson
Associate Professor of Educational Leadership and Research
University of Southern Mississippi
Southern Station Box 9221
Hattiesburg, Mississippi 39401

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

W. F. Ferguson

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Descriptors: Answer-Sheet Format
Cheating
Testing

TM 830 809

Multiple-choice tests have the reputation among many of the so-called "experts" of being the best type of objectively-scored test. This does not mean, of course, that multiple-choice tests do not have their faults or their critics. Hoffman (1961), in a scathing article, referred to the "tyranny of multiple-choice tests," but even he begrudgingly admitted the value of that test format before ending his article. Like it or not, this test format is with us and will continue to be a primary force in decision-making for class grades, for admission to college, for admission to professional schools, and for other such purposes for a long time to come.

Instructors, whether at the college level or below, often find themselves with more students in their classes than they can adequately supervise during examination/test times. The problem is particularly critical when instructors utilize objectively-scored test items as even a portion of a complete test grade. Cheating on an objectively-scored test is far easier than is cheating on a subjectively-scored test, particularly on an essay-type test. In spite of this potential problem, many instructors prefer and continue to use the objectively-scored test format while trying to lessen opportunities for students' copying from other students' papers without their getting caught. Occasionally, generally at the college level, larger rooms can be located which are more conducive to assuring the validity of each student's score. Perhaps another way to

remedy the situation of the possibility of cheating is with the addition of one or more "evil-eyed" proctors. When these possibilities are not available, instructors often turn to multiple forms of the same test (i.e., forms A and B). Sometimes forms A and B are composed of entirely different items and sometimes they are composed of the very same items in a different order or even in a different form (i.e., true-false versus multiple-choice, or some other such combinations.)

While any one of the above possibilities probably could be used to either solve or alleviate the problem of students cheating on tests, each one has the outstanding disadvantage of causing extra work for the instructor and/or the typist. One possible solution (Disney, Merrifield, and Davis; Miller and Minor; Slater; and Whitcomb) which has had very little exposure is to create only one form of the examination but utilize two different forms of the response/answer sheet. From this possible solution, the question then should naturally arise, "Is a student with a non-traditional answer sheet at a disadvantage to a student with a traditional answer sheet?"

SUBJECTS: Subjects consisted initially of 42 students (primarily college juniors and seniors) enrolled in REF 400--Public Education in the United States during the fall semester of 1982 at the University of Southern Mississippi's Hattiesburg campus. Four students dropped prior to the final test; data analysis, consequently, utilized the responses of 38 to 42 students on various occasions. Additionally, for the data

analysis for "paired scores," requiring the exclusion of all students for whom both scores were not available, the number of usable pairs of scores occasionally fell to as low as 35 pairs.

The 42 students were assigned to meet the class in a classroom adequate for testing no more than 25 students under good testing conditions. Consequently, the students were seated shoulder to shoulder, ten per horizontal row over four rows of tables, with the two extras seated in arm-chair desks along a side wall of the classroom. Obviously an instructor who administered multiple-choice items on his tests would be inviting cheating under normal conditions. What solution would offer the most test security at the least effort?

INSTRUMENTATION: It was decided that the instructor would try a different solution from the usual one of multiple forms of the same test. A possible solution of utilizing the same test with different response sheets was considered, but no research could be located which utilized multiple answer sheets for this purpose. Then, too, what few research studies could be found, did not offer formats that the instructor felt to be both sensible and difficult to decipher at a glance from a short distance. Finally, it was decided to try one answer form in the traditional fashion--items numbered vertically and alternative choices listed horizontally. The second form was to also have the alternative choices listed horizontally, but the items themselves were to be listed by threes horizontally (i.e., the first vertical column would contain items 1, 4, 7, 10, etc.;

the second vertical column would contain items 2, 5, 8, etc.; and the third column would contain items 3, 6, 9, etc. Other than that variation, the answer sheets were to be virtually identical.

METHODOLOGY: Students were not appraised of the fact that they would have identical tests with different answer sheets until they were seated according to a randomly-devised seating chart. Numbered tests were passed out according to a pre-conceived plan, and then the answer sheets were passed out. Students were instructed to write their names and the test numbers on their answer sheets. The test numbers served the dual purpose of keeping each student's test and answer sheet together and also of allowing comparisons of adjacent answer sheets in the event of suspected cheating. Students who received the vertically numbered answer sheet on the first test were to receive the horizontally numbered answer sheet on the next test.

Students were very casually forewarned prior to the administration of test 1A that they "might" get a second test over the same material if the instructor was not satisfied with their scores on the first test. Nothing else was said by either instructor or students about the possibility of a second test prior to it being administered on the next scheduled class meeting following the one when test 1A was given, other than only the highest of the two grades would count toward their semester average and that absence on the scheduled test date (unless totally beyond their control) would

cause their best grade of the two to be their only grade. That is to say that an invalid excuse on one of the two test days would automatically cause a default of only one grade to be counted over that material.

Of the originally scheduled four tests during the semester, tests one and three were retested with entirely different test questions, also multiple-choice, resulting in scores from tests 1A, 1B, 2, 3A, 3B, and 4. Again, test 1B was merely stated as a "possibility." Before test 2 was administered, students asked the instructor about the possibility of a test 2A and a 2B; the instructor's responses to this question were purposefully vague and noncommittal. However, subjects were told in very specific terms before test 3A was administered that there would be a second test over that material. Then, for the final examination, students were told very definitely that test 4 would be given only once, regardless of their grades or any other factors. Therefore, each of the four major content divisions of the course had tests administered (or not administered) under four completely different instructions. These conditions also seemed to play a role in the final outcome of the students' scores. Additionally, while a conscious effort was made by the instructor to assure that forms A and B of tests 1 and 3 were equivalent, no significant attempt was made to assure the equivalence of tests 1, 2, 3, and 4.

FINDINGS: Mean scores on the six separate tests (1A, 1B, 2, 3A, 3B, and 4) were 74.83, 72.28, 75.60, 81.32, 85.37, and

85.00 respectively. In no case was there a statistically significant difference between the means on each test when the scores were examined by type of answer sheet utilized. In fact, none of the differences in means even began to approach significance; however, there were very persistent but non-significant differences in their resulting standard deviations. (See Table 1 for a representative sample of this phenomenon.) Table 2 is a graphic representation of the distribution of grades made by the students who were present for the administration of tests 1A and 1B. Each student has a grade recorded that was earned on test 1 by both vertically numbered and horizontally numbered answer sheets. This chart or table typlifies all such resulting grade distributions in this study just as Table 1 was typical of tables similar to it.

Perhaps the most interesting findings emerged when the scores on the vertically and horizontally numbered answer sheets were examined in light of the students' final letter grades in the course. For those students whose final average was an A, it was noted that their average grade on the vertical answer sheets was approximately two points higher than their average grade on the horizontally numbered answer sheets. For students who earned a letter grade average in the course of B, their vertically numbered answer sheet average was approximately one-half point greater than their horizontally numbered answer sheet average. For those who earned a letter grade of C, their vertically numbered answer sheet average exceeded their horizontally numbered answer sheet average by

roughly one point. However, for those students whose final grade in the course was a D, their average on the vertically numbered answer sheets was approximately four points lower than their average on the horizontally numbered answer sheets. And, for those who ended up with an F (including those who "dropped" failing but who did not actually receive a letter grade of F), their vertically numbered answer sheet average was more than six and one-half points below that of their horizontally numbered answer sheet average.

CONCLUSIONS: There are two types of conclusions which can be drawn as a result of an analysis of the complete data gathered throughout the entire semester. The first type is applicable to all such studies and the other type is applicable specifically to this study. The first type will be examined first. Perhaps as much as anything else this study points out and/or reinforces the following concepts:

(1) When students are tested a second time (without advance notice the second time only) over the same material, their second mean score will be lower than their first mean score. This phenomenon reflects the "memory [forgetting] curve."

(2) When students are to be tested only once over a specified set of content and expect a second test (although no second test was promised) to also be given, their preparation (as reflected by their scores and reinforced by their later comments) for the second test is not as well done as when they expect only one test

over that same content. This conclusion must be tempered with the fact that the subjects of this study knew in advance that only the higher of the two scores--in the event of a second test--would count toward their semester average.

The second type of conclusion deals specifically with the present study. Although the response sheet formats gave no significant differences in mean scores between groups, there was a consistent pattern of a reduction of variance in the scores of the horizontally numbered answer sheet group. Additionally, students who performed best in the course academically, as reflected in their overall letter grade, did slightly better on the vertically numbered forms. Conversely, students who did the worst academically actually did considerably better on the horizontally numbered answer sheet than on the vertically numbered one. This, along with personal observation over all testing situations, leads the researcher to conclude that the non-traditional (horizontal) answer sheet format was somewhat confusing to the students and perhaps "forced" them to reread each test item and recorded answer in order to be sure that they recorded the answer they intended in the place they intended.

SELECTED BIBLIOGRAPHY

Cates, W. M. "The Efficacy of Retesting in Relation to Improved Test Performance of College Undergraduates," Journal of Educational Research, 75:230-236, March/April, 1982.

Dizney, H. F., P. R. Merrifield, and O. L. Davis, Jr. "Effects of Answer-Sheet Format on Arithmetic Test Scores," Educational and Psychological Measurement, 26:491-493, Summer, 1966.

Duchastel, P. C. and R. J. Nungester. "Testing Effects Measured with Alternate Forms," Journal of Educational Research, 75:303-313, May/June, 1982.

Green, K. "Item-Response Changes on Multiple-Choice Tests as a function of Test Anxiety," Journal of Experimental Education, 49:225-228, Summer, 1981.

Hoffman, B. "The Tyranny of Multiple-Choice Tests," Harper's Magazine, 222:37-44, March, 1961.

Howard, K. I. "Differentiation of Individuals as a function of Repeated Testing," Educational and Psychological Measurement, 24:875-894, Winter, 1964.

Miller, I. "A Note on the Evaluation of a New Answer Form," Journal of Applied Psychology, 49:199-201, June, 1965.

Miller, I. and F. J. Minor. "Influence of Multiple-Choice Answer Form Design on Answer Marking Performance," Journal of Applied Psychology, 47:374-379, December, 1963.

Plake, B. S. "Item Arrangement and Knowledge of Arrangement on Test Scores," Journal of Experimental Education, 49:34-37, Fall, 1980.

Schrock, T. J. "Effects of Violating Three Multiple-Choice Item Construction Principles," Journal of Educational Research, 75:314-318, May/June, 1982.

Slater, R. D. "The Equivalency of IBM Mark-Sense Answer Cards and IBM Answer Sheets when Used as Answer Formats for a Precisely-Timed Test of Mental Ability," Journal of Educational Research, 57:545-547, July-August, 1964.

Whitcomb, M. A. "The IBM Answer Sheet as a Major Source of Variance on Highly Speeded Tests," Educational and Psychological Measurement, 18:757-759, 1958.

TABLE 1

ANALYSIS OF VARIANCE
1-A VERTICAL VS. 1-A HORIZONTAL

GROUP	N	S.D.	MEAN	F	p
1-A VERTICAL	21	20.2738	75.1429	.01275	.91
1-A HORIZONTAL	20	14.0955	74.5000		

TABLE 2

PERCENTILE ANALYSIS
REF 400 1:A+B VERTICAL VS. HORIZONTAL
(N = 38 pairs)

VERTICAL		SCORES	HORIZONTAL	
%ile	f		f	%ile
97.3	2	100		
		99		
		98		
		97		
91.2	2	96	1	98.6
		95		
84.2	4	94		
		93		
77.6	1	92	3	93.4
		91		
75.0	1	90	4	84.2
		89		
72.3		88	2	76.3
		87		
69.7	1	86		
		85		
65.7	2	84	4	68.4
		83		
61.8	1	82	1	61.8
		81		
56.5	3	80	3	56.5
		79		
51.3	1	78	2	50.0
		77		
47.3	2	76	1	46.0
		75		
40.7	3	74		
34.2	2	73		
30.2	1	72	2	42.1
		71		
		70		
		69		
		68	1	38.1
26.3	2	67	3	32.8
		66	1	27.6
22.3	1	65	1	25.0
		64	1	22.3
		63		
		62		
		61	1	19.7
		60	1	17.1

(Continued on right side)

(Continued)

%ile	f		f	%ile
		59		
		58		
		57		
19.7	1	56		
		55		
		54		
17.1	1	53	1	14.4
		52		
14.4	1	51	1	11.8
11.8	1	50		
		49		
		48		
		47	1	9.2
		46	1	6.5
		45		
		44	1	3.9
		43		
		42		
		41		
		40		
		39		
		38		
9.2	1	37		
		36		
		35	1	1.3
6.5	1	34		
		33		
		32		
		31		
		30		
3.9	1	29		
1.3	1	28		